

Сокращение размерности данных в задачах имитационного моделирования

Ю.Г. Агалаков, А.В. Бернштейн

Аннотация. Рассматриваются задачи интеллектуального анализа данных, которые необходимо решать в технологии предсказательного моделирования. Для уменьшения сложности решения этих задач в технологии предсказательного моделирования используются решения задач снижения размерности, которые должны удовлетворять ряду дополнительных условий. В статье обсуждаются эти дополнительные требования и сформулированы соответствующие новые нетрадиционные постановки задач снижения размерности.

1. Введение. Технология предсказательного моделирования.

В последние годы компьютерное проектирование стало необходимым (а зачастую и важнейшим) элементом в различных областях, таких как авиакосмическая промышленность, информационные управляющие системы, телекоммуникационные системы, биотехнология, нефтегазовый сектор и многих других. В процессе проектирования всегда происходит сравнение различных технических решений, касающихся структуры, параметров и механизмов функционирования проектируемого объекта (называемых далее параметрами проектирования или инструментальными переменными), исходя из их влияния на характеристики поведения объекта в тех или иных условиях функционирования и при различных управляющих воздействиях. Например, в аэрокосмической промышленности характеристиками проектируемого летательного аппарата могут быть аэродинамические и прочностные параметры, а в информационных управляющих и телекоммуникационных системах в качестве таковых выступают вероятностно-временные характеристики.

Явные аналитические зависимости характеристик от параметров проектирования, условий функционирования и управляющих воздействий невозможно получить для сколько-нибудь сложного реального объекта проектирования. Натурные испытания могут быть, как правило, проведены лишь на заключительных этапах проектирования, когда объект уже создан и может эксплуатироваться, а основные технические решения уже приняты и их изменения уже невозможны или сопряжены с большими затратами. Поэтому в процессе проектирования все чаще применяются методы компьютерного моделирования, построенные на вычислительных экспериментах с математическими моделями проектируемого объекта и окружающей его среды. Компьютерное моделирование применяется также для сокращения количества дорогостоящих (в материальном и временном смысле) натурных экспериментов. Для того, чтобы подчеркнуть, что целью компьютерного моделирования является предсказание (прогноз,

оценка) характеристик проектируемого объекта, в последнее время часто используется термин «предсказательное моделирование».

Компьютерные системы предсказательного моделирования (называемые также системами поддержки принятия инженерных решений) вместе с компьютерными системами проектирования давно используются для автоматизации труда инженера-проектировщика и повышения качества принимаемых решений. Однако до начала 21 века в предсказательном моделировании использовались исключительно математические модели, основанные на "первых принципах физики", описывающие физические процессы и явления, происходящие при функционировании объекта, сложными дифференциальными уравнениями в частных производных с граничными условиями. В сколько-нибудь содержательных ситуациях для этих уравнений неизвестны ни теоремы о существовании и единственности решения, ни характер зависимости решения от параметров и граничных условий. Численные методы решения таких уравнений имеют значительную вычислительную трудоемкость как самих расчетов, так и подготовки исходных данных и расчетных сеток. Это существенно сокращает возможности использования таких моделей в проектировании сложных объектов, особенно на стадии предварительного (концептуального) проектирования, когда рассматривается большое количество разнообразных вариантов решений и особенно высока цена неправильно выбранного решения.

Важной частью предсказательного моделирования является имитационное моделирование, используемое для исследования сложных информационно-телекоммуникационных систем (ИТКС). Для проектирования и исследования информационно-вычислительных сетей (ИВС), являющихся телекоммуникационными компонентами ИТКС, имитационное моделирование является мощным инструментом, позволяющим предсказывать вероятностно-временные и надежность характеристики сетей, проектировать и исследовать процедуры управления сетью. Например, при имитационном моделировании ИВС, состоящей из узлов коммутации, связанных между собой трактами передачи данных, состоящими в общем случае из каналов разной физической природы [15], в каждом такте моделирования необходимо имитировать передачу информации по каждому тракту и обработку информации в каждом узле коммутации. Эти процессы передачи и обработки информации в компонентах ИВС описываются сложными системами массового обслуживания, входные потоки на которые, в силу наличия в сети процедур управления потоками, зависят от загрузки всех трактов и узлов коммутации. Для моделирования этих частных процессов создаются частные достаточно сложные и ресурсоемкие имитационные модели, одновременное использование которых для всех трактов

и узлов делает время моделирования неприемлемо высоким по сравнению с временем работы реальной системы. В смешанных системах моделирования сначала подробно моделируют тракты и узлы ИВС, а по результатам моделирования оценивают функции распределения времен задержек информации, с помощью которых имитируются задержки информации в различных компонентах сети, однако так построенные функции распределения времен задержки не дают адекватного описания сложных процессов передачи и обработки информации. Кроме того, для каждого нового набора входных данных (матрицы интенсивностей входных потоков, параметров трактов передачи данных и т.п.) надо проводить новые имитационные эксперименты, требующие достаточно много времени, для получения статистически достоверных оценок для требуемых характеристик. Поэтому имитационное моделирование также является «дорогостоящим» инструментом исследования, что также сокращает возможности использования имитационного моделирования в проектировании сложных ИТКС, особенно на стадии проектирования.

Как ответ на возникшие сложности, в последние годы стало активно развиваться новое, альтернативное направление математического моделирования - метамоделирование, в рамках которого математические предсказательные модели для сложных систем строятся по результатам вычислительных экспериментов с моделями, основанными на первых принципах физики, или с имитационными моделями. Построение таких моделей основано на идеях машинного обучения (machine learning) [9 - 11], где модели "обучаются" по множеству прототипов входных и выходных данных – результатах экспериментов. Построенные модели фактически имитируют (заменяют) как источники получения данных, основанные на некоторой исходной модели, так и сами модели, созданные на основе изучения физики процессов. Поэтому такие адаптивные модели иногда называют также метамоделями (модели над моделями) или суррогатными моделями (Surrogate Models). Как правило, метамодели имеют существенно более высокую вычислительную эффективность по сравнению с исходными моделями (иначе в них просто нет смысла). Например, метамодели для расчета аэродинамических характеристик, построенные по результатам экспериментов с исходной CFD-моделью, основанной на численном решении дифференциальных уравнений аэродинамики, увеличили скорость вычислений более чем в сотни тысяч раз при относительной погрешности не более 1% [12 - 14].

Для построения метамodelей необходимо решать ряд взаимосвязанных задач интеллектуального анализа данных, в которых выходные данные одной частной задачи являются входными данными для другой задачи, и целевые функции для частных задач нельзя

определить независимо [16 - 23]. Формальные постановки решаемых задач имеют ряд особенностей, отличающих их от классических постановок. Например, носители обрабатываемых многомерных данных обычно лежат на нелинейных многообразиях меньшей размерности (многообразиях данных), что не учитывается в стандартных математических методах анализа данных [24 - 26]. Прежде всего, на практике очень часто приходится иметь дело с объектами, описываемыми огромным числом параметров, для анализа которых многие процедуры обработки данных оказываются неэффективными или просто неработоспособными. В этом случае возникает естественный вопрос, а нельзя ли попытаться каким-то образом сократить размерность параметрического описания объектов без потери точности модели.

В статье предложены формальные постановки ключевых задач интеллектуального анализа, возникающих в технологии предсказательного моделирования. Подробно обсуждены постановки нестандартных задач снижения размерности.

2. Интеллектуальный анализ данных в предсказательном метамоделировании

В самом общем виде базовая задача построения метамоделей формулируется следующим образом. Пусть $\mathbf{O} = \{O\}$ есть множество объектов рассматриваемого класса. Для каждого объекта $O \in \mathbf{O}$ имеется его цифровое описание $X = X(O)$ размерности p , полностью определяющее этот объект. Обозначим

$$\mathbf{X} = X(\mathbf{O}) = \{X(O), O \in \mathbf{O}\} \subset \mathbb{R}^p \quad (1)$$

подмножество p -мерного евклидова пространства, состоящее из цифровых описаний объектов рассматриваемого класса.

Обозначим $Z = Z(O)$ некоторую, в общем случае, векторную, характеристику объекта O , описывающую свойство (поведение) объекта в некоторых условиях. Характеристика Z может быть описана в виде функциональной зависимости характеристики

$$Z = F(X), \quad X \in \mathbf{X}, \quad (2)$$

от цифрового описания $X = X(O)$ объекта O .

Для вычисления значений функции $Z = F(X)$ могут использоваться различные методы, например, вычислительные или натурные эксперименты. Пусть M – некоторый метод (модель) получения характеристики Z , тогда модель M может быть представлена в виде зависимости

$$Z = F_M(X), \quad X \in \mathbf{X}, \quad (3)$$

где $F_M(X)$ есть значение характеристики Z для объекта с цифровым описанием X , полученное с помощью модели M и являющееся приближенным значением $F(X)$ характеристики Z (2):

$$F_M(X) \approx F(X) \quad \text{для всех } X \in \mathbf{X}. \quad (4)$$

Пусть имеется некоторое количество n измерений

$$\mathbf{D} = \mathbf{D}_n = \mathbf{D}_n(\mathbf{M}) = \{(X_i, Z_i = F_M(X_i)), i = 1, 2, \dots, n\}, \quad (5)$$

характеристики Z , полученных с помощью модели M для разных значений аргументов, принадлежащих множеству

$$\mathbf{X}_n = \{X_i, i = 1, 2, \dots, n\} \subset \mathbf{X}. \quad (6)$$

Предполагается, что модель M имеет приемлемую точность (выполняется условие (4)).

Множество \mathbf{D}_n (5) состоит из известных значений неизвестной функции $F_M(X)$ в точках множества \mathbf{X}_n (6), по которым может быть построена эмпирическая зависимость

$$Z = F_{SM}(X) \equiv F_{SM}(X|\mathbf{D}_n), \quad (7)$$

аппроксимирующая функцию $F_M(X)$. Если в области \mathbf{X} изменения переменных X имеют место приближенные равенства

$$F_{SM}(X|\mathbf{D}_n) \approx F_M(X) \quad \text{для всех } X \in \mathbf{X}, \quad (8)$$

то эмпирическая функция $F_{SM}(X)$ определяет новую модель (метамодель) SM , которая может рассматриваться как заменитель (surrogate) для модели M . Как правило, функция $F_{SM}(X)$ имеет явный, хотя часто и сложно записываемый, аналитический вид, и поэтому метамодель SM может иметь существенно более высокую вычислительную эффективность по сравнению с исходной моделью M .

Ключевой задачей интеллектуального анализа данных в метамоделировании является задача построения по данным (5) аппроксимационной зависимости $F_{SM}(X|\mathbf{D}_n)$ (7). Величину погрешности в (8) для точек множества \mathbf{X}_n (6) можно непосредственно вычислить и проконтролировать, а погрешность для новых (out-of-sample) точек $X_{new} \in \mathbf{X} / \mathbf{X}_n$ характеризует обобщающую способность построенной метамодели $F_{SM}(X)$. Обзоры методов параметрической и непараметрической регрессии для построения аппроксимационных зависимостей, обеспечивающих необходимую обобщающую способность процедуры, можно найти в работах [2, 4, 6 - 8, 16, 20]. К этой задаче тесно примыкают задачи построения по данным (5) процедур оценивания различных функционалов от величины ошибки

$$\delta_{SM}(X) = F_{SM}(X) - F_M(X), \quad (9)$$

в том числе задачи построения доверительных интервалов для функции $F_M(X)$.

Построенная суррогатная модель $F_{SM}(X|\mathbf{D}_n)$ в силу очевидного неравенства

$$|F_{SM}(X) - F(X)| \leq |\delta_M(X)| + |\delta_{SM}(X)|$$

и соотношений (4) и (8) может служить приближением к «истинной» функции $F(X)$, здесь

$$\delta_M(X) = F_M(X) - F(X) \quad (10)$$

есть погрешность модели M , результаты \mathbf{D}_n экспериментов с которой использованы для построения метамодели SM .

Величина $\delta_{SM}(X)$ зависит от размера n обучающего множества D_n . Очевидно, что стоимость (в разных ресурсных смыслах) экспериментов зависит от модели M – достаточно сравнить натурные эксперименты (например, натурные эксперименты в аэродинамических трубах) и вычислительные эксперименты с математическими моделями (например, расчеты с помощью CFD-кода). Стоимость вычислительных экспериментов, в свою очередь, существенно зависят от сложности выбранной математической модели – например, время аэродинамических расчетов по моделям, основанным на уравнениях Навье-Стокса, превышает в сотни и тысячи раз время расчетов на более простых моделях, основанных на уравнениях Эйлера или полного потенциала. В то же время более «дорогие» модели имеют, как правило, и более высокую точность (меньшие значения погрешности $\delta_M(X)$ (10)). Поэтому в реальных приложениях трудно иметь большие размеры обучающих выборок, полученных в экспериментах с дорогими «высокоточными» моделями, но можно провести достаточно много экспериментов с «дешевыми», но «низкоточными» моделями. Поэтому в технологии метамоделирования возникает следующая задача интеллектуального анализа данных.

Предположим, что для дешевой низкоточной модели M , являющейся приближением «истинной» функции $F(X)$, построена метамодель $F_{SM}(X)$, которая в силу доступности больших обучающих выборок обеспечивает достаточно малую погрешность $\delta_{SM}(X)$ (9). Пусть имеется другая, более дорогая, модель M^* , также являющаяся приближением той же самой «истинной» функции $F(X)$, но с существенно меньшей погрешностью $\delta_{M^*}(X)$ (10). Предполагается, что количество экспериментов n^* , которые можно провести с моделью M^* , явно недостаточно для построения метамодели SM^* с приемлемой величиной погрешности $\delta_{SM^*}(X)$ (9). Ставится задача построения функции

$$F^*(X) = F^*(X|D_{n^*}, F_{SM}(\cdot)) \quad (11)$$

по результатам $D_{n^*}(M^*)$ экспериментов с моделью M^* и с использованием метамодели SM .

Простейшим вариантом (9) является функция вида

$$F^*(X) = F_{SM}(X) + \Delta_{SM,M^*}(X),$$

где $\Delta_{SM,M^*}(X)$ – метамодель для функции

$$\Delta_{SM,M^*}(X) = F_{M^*}(X) - F_M(X),$$

построенная по наблюдениям $\{(X_i, F_M(X_i), F_{M^*}(X_i)), i = 1, 2, \dots, n^*\}$, в которой множество результатов экспериментов $D_{n^*}(M^*)$ с моделью M^* дополнено результатами экспериментов с моделью M в точках, в которых были проведены эксперименты с моделью M^* . Вычислительные эксперименты с реальными и искусственными данными [27 - 29] показали, что число экспериментов n^* с моделью M^* существенно меньше числа экспериментов, которые

необходимо было бы провести для построению метамоделей SM^* , не использующей значения метамоделей $F_{SM}(X)$, для достижения той же точности, что и модель (11).

3. Снижение размерности данных в предсказательном моделировании

3.1. Содержательная постановка проблемы снижения размерности

В приложениях размерность p цифрового описания объекта может быть очень высока. Например, детальные описания геометрических объектов (кривых, поверхностей) или их компонентов в общем случае задаются набором 2D- или 3D-координат точек поверхности объекта, лежащих в выбранных узлах геометрического объекта. Такие цифровые описания объекта (например, 3D-поверхности самолета) состоят из десятков тысяч чисел, каждое из которых, рассматриваемое изолированно, не несет смысловой нагрузки [25].

Другим примером высокой размерности входных данных являются модели, описывающие процессы передачи и обработки информации в ИВС. Рассмотрим подробнее проблему построения метамоделей для передачи информации по тракту передачи данных. Информационно-вычислительную сеть можно представить в виде графа (Рис. 1), узлы которого являются центрами коммутации, а ребра – трактами передачи данных [15].

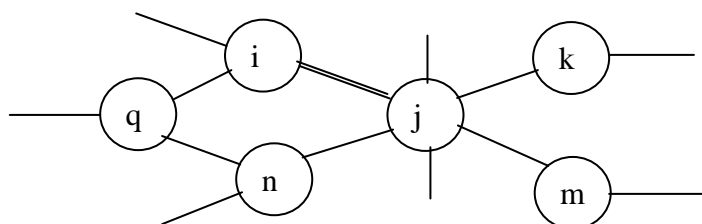


Рис. 1. Фрагмент графа, описывающего ИВС

Пусть в ИВС циркулирует информация разных приоритетов, и информация более высокого приоритета «обгоняет» информацию более низкого приоритета при передаче по трактам передачи данных. Поэтому процесс передачи по тракту можно достаточно адекватно описать моделями систем массового обслуживания с приоритетами, достаточно изученными в литературе [30 - 33].

Пусть $X_{jk}(s, t)$ - длина очереди сообщений s -го приоритета, в момент времени t для передачи по тракту передачи данных (j, k) . Однако время задержки $T_{jk}(S, t)$ сообщения с приоритетом S , ставшего в момент времени t в эту очередь, определяется не только количеством сообщений $X_{jk}(s, t)$, $s \leq S$, уже стоящих в очереди, но и количеством сообщений $X_{ij}(s, t)$, $X_{nj}(s, t)$, $X_{mj}(s, t)$, $X_{qn}(s, t)$, ... (см. Рис. 1), $s < S$, которые стоят в очередях на передачу в другие тракты передачи данных (те из них, которые будут далее передаваться по тракту (j, k) и, в силу более высокого

приоритета, потенциально могут обогнать и, следовательно, дополнительно задержать рассматриваемое сообщение).

В силу наличия в сети процедур управления потоками, выбор направления дальнейшей передачи сообщений производится адаптивно, исходя из текущей загрузки всех трактов и узлов коммутации. Поэтому входной поток в тракт передачи данных зависит от состояния приборов массового обслуживания, и аналитическая оценка величины $T_{jk}(S, t)$ существующими методами невозможна. Если мы будем строить метамодель для величины $T_{jk}(S, t)$ для разных приоритетов, то входной вектор для нее будет включать очереди в различные тракты передачи данных по различным приоритетам, то есть иметь достаточно высокую размерность.

Высокая размерность цифрового описания объектов существенно затрудняет аппроксимацию нелинейной функции, зависящей от векторов высокой размерности, так как требуется выборка достаточно большого объема n для достижения приемлемой точности. Однако даже такие высокоэффективные алгоритмы как ядерная гребневая регрессия (кригинг) [34 - 37] или эффективные реализации нейросетевых алгоритмов с использованием алгоритмов оптимизации Левенберга-Маркварта [38] быстро теряют свою эффективность с ростом чисел p и n , или просто становятся неработоспособными.

К счастью, в приложениях, как правило, в силу многочисленных связей между компонентами цифрового описания объекта и ограничений на них, множество X лежит, по крайней мере приближенно, на многообразии (в общем случае, нелинейном) M , размерность q которого существенно меньше исходной размерности p . Например, как показано в работах [12, 14, 39, 40], множество X , состоящее из описаний аэродинамических профилей и лежащее в пространствах размерности $50 \div 100$ (в зависимости от требуемой точности цифрового описания) [41], может быть аппроксимировано многообразием размерности $5 \div 10$.

Если удастся аппроксимировать множество X многообразием M меньшей размерности, то задачу аппроксимации необходимо решать лишь в окрестности многообразия M . Нахождение такого многообразия M очень важно и по другой причине. Для проведения вычислительных экспериментов с целью получения обучающего множества данных D , необходимо генерировать входные цифровые описания X в окрестности многообразия M , а в итерационных процессах оптимизации также необходимо «оставаться» вблизи многообразия M .

3.2. Формальная постановка задачи снижения размерности

Задача построения аппроксимирующего многообразия M может быть сформулирована как Задача снижения размерности, или, в современной терминологии, как Задача моделирования многообразий (Manifold Learning Problem). Будем, как и выше, рассматривать множество

объектов $\mathbf{O} = \{O\}$, цифровое описание каждого из которых является p -мерным вектором $X = X(O) \in \mathbb{R}^p$, называемым детальным (высоко размерным, полноразмерным, полным) описанием объекта O . Пусть $\mathbf{X} \subset \mathbb{R}^p$ (1) - множество детальных описаний всех объектов из класса \mathbf{O} .

Снижение размерности заключается в построении для каждого объекта $O \in \mathbf{O}$, заданного своим детальным описанием $X(O)$, вектора $y = y(O) \in \mathbb{R}^q$ меньшей размерности $q < p$, называемого сокращенным (низко размерным, сжатым) описанием объекта O и содержащего в себе необходимую информацию об объекте.

Предполагается, что множество \mathbf{O} известно только множеством прототипов $\mathbf{O}_n = \{O_1, O_2, \dots, O_n\} \subset \mathbf{O}$, определяющим множество их детальных описаний

$$\mathbf{X}_n = \{X_1, X_2, \dots, X_n\}, \quad X_i = X(O_i), \quad i = 1, 2, \dots, n, \quad (12)$$

а сокращенные описания объектов должны строиться только на основе множества данных \mathbf{X}_n , без привлечения знаний из предметной области.

Имеется несколько различных постановок задач снижения размерности, основные из которых приведены ниже. В литературе для этих задач используются различные наименования, и используемая ниже терминология не претендует на всеобщее использование.

Задача вложения. По заданному множеству $\mathbf{X}_n \subset \mathbb{R}^p$ (12) необходимо построить отображение

$$h: \mathbf{X} \subset \mathbb{R}^p \rightarrow \mathbf{Y} = h(\mathbf{X}) \subset \mathbb{R}^q, \quad (13)$$

сохраняющее заданные соотношения между точками множеств \mathbf{X} и $h(\mathbf{X})$. Обычно от отображения h требуют, чтобы точки множества \mathbf{Y} сохраняли геометрическую структуру точек множества \mathbf{X} (локальную геометрию, отношения близости, геодезические расстояния).

Иногда задача вложения формулируется только для точек множества \mathbf{X}_n , и необходимо построить отображение

$$h: \mathbf{X}_n \rightarrow \mathbf{Y}_n = h(\mathbf{X}_n) = \{y_1, y_2, \dots, y_n\} \subset \mathbb{R}^q, \quad (14)$$

множества \mathbf{X}_n в множество точек \mathbf{Y}_n , лежащих в пространстве меньшей размерности $q < p$, с сохранением структуры множества \mathbf{X}_n . Полученные точки $y_i = h(X_i) \in \mathbf{Y}_n$, рассматриваемые как малоразмерные представления многомерных векторов $X_i \in \mathbf{X}_n$, должны «достоверно представлять» выборку \mathbf{X}_n .

Отображение h (13), (14) зависит от выбранных свойств, которые должны сохраняться при переходе от детальным к сокращенным описаниям. Например, для целей визуализации множество многомерных данных \mathbf{X}_n отображается в множество \mathbf{Y}_n , лежащее в пространстве размерности $q = 2$ или 3 , и должна сохраняться близость между соответствующими точками в множествах \mathbf{X}_n и \mathbf{Y}_n . Для решения этой задачи может быть использован метод Многомерного шкалирования (Multi Dimensional Scaling) [42], в котором отображение h (14) строится путем

минимизации функционала $W(Y_n) = \sum_{i,j} (\rho(X_i, X_j) - \|y_i - y_j\|)^2$, где ρ - выбранная метрика в пространстве исходных многомерных векторов. В случае евклидовой метрики ρ метод Многомерного шкалирования эквивалентен широко известному Методу главных компонент (Principal Component Analysis) [43, 44]. В других прикладных областях, таких как распознавание лиц, голоса, символов текста [45 - 48], ищутся описания $y(O)$ невысокой размерности, которые можно использовать вместо детальных высоко размерных описаний $X(O)$ при решении задачах классификации, распознавания и т.п. В такой постановке задача вложения рассматривалась в работах [49 - 56] и др., в которых предложены различные подходы к ее решению.

Если имеется некоторый метод, позволяющий решать Задачу вложения (14) для произвольного множества точек X_n , то его формально можно использовать и в Задаче вложения (13) для произвольной новой (out-of-sample) точки $X_{new} \in X / X_n$, заново решая задачу для множества $\{X_n \cup X_{new}\}$. Но при этом в общем случае не обеспечивается сохранение результатов вложения $Y_n = h(X_n)$ (14), ранее полученного для исходной выборки X_n . Построение вложения h для точек множества $\{X_n \cup X_{new}\}$, сохраняющего множество $Y_n = h(X_n)$, рассматривалось в работах [57 - 59] и др., и метод работ [57, 58] назван Методом расширения вложения для новых точек (out-of-sample extension). В соответствии с этим определением, задачу построения вложения h (13) для всех точек множества X , в отличие от задачи построения вложения h (14) только для точек выборки X_n , будем называть **Расширенной задачей вложения**.

В общем случае при переходе от исходного детального описания $X(O) \in R^p$ объекта O к сокращенному описанию $y(O) = h(X(O)) \in R^q$ меньшей размерности $q < p$ может происходить значимая потеря информации об объекте O , если от вложения h (13) требуется лишь сохранение конкретных вполне определенных свойств объекта. В прикладных задачах часто требуется, чтобы при использовании сокращенного q -мерного описания вместо детального p -мерного описания не происходило значимой потери информации об объекте, понимая под этим, что по сокращенному описанию можно приближенно восстановить детальное описание объекта. Будем называть такую задачу **Полной задачей снижения размерности**.

Полная задача снижения размерности. По выборке X_n (12) из множества детальных описаний $X(O)$ необходимо построить отображение вложения h (11) и отображение восстановления

$$g: Y \subset R^q \rightarrow X^* = g(Y) \subset R^p, \quad (15)$$

восстанавливающее детальное описание объекта $X(O)$ по его сокращенному описанию $y(O)$, такие, что пара отображений $\theta = (h, g)$ должна обеспечивать близость

$$X \approx g(h(X)) \quad \text{для всех } X \in \mathbf{X} \quad (16)$$

между исходными детальными описаниями \mathbf{X} и восстановленными детальными описаниями $g(h(X))$, являющимися результатом последовательного применения отображений вложения и восстановления.

Заметим, что если требовать выполнения свойства (16) только для точек выборки \mathbf{X}_n , то полная задача снижения размерности, без дополнительных требований на ее решение $\theta = (h, g)$, становится вырожденной и бессмысленной – можно построить процедуры вложения и восстановления (14), (15), обеспечивающие безошибочное восстановление даже для одномерных ($q = 1$) сокращенных описаний. Например, достаточно взять в качестве вложения h (14) точки $X_i \in \mathbf{X}_n$ ее номер i ($h(X_i) = i$), а в качестве отображения восстановления g (15) функцию $g(y) = X_y$ для $y \in \mathbf{Y}_n = h(\mathbf{X}_n) = \{1, 2, \dots, n\}$.

Этот вырожденный пример отражает очень важный феномен в проблеме снижения размерности. В Расширенной задаче вложения, так же, как и в полной задаче снижения размерности, необходимо определить значения функции h (13) для новых точек $X \in \mathbf{X} / \mathbf{X}_n$ и значения функции g (15) для точек $y = h(X) \notin \mathbf{Y}_n$, являющихся образами новых точек после операции вложения. Для произвольного решения $\theta = (h, g)$ можно «контролировать» его качество (выполнение приближенных соотношений (16)) для точек выборки, но при этом решение θ может иметь сколь угодно низкую обобщающую способность, под которой в теории машинного обучения [9 – 11] применительно к рассматриваемой задаче понимается выполнение соотношений (16) для точек $X \in \mathbf{X} / \mathbf{X}_n$. Но для построения решения θ , обладающего обобщающей способностью, необходимо сделать некоторые предположения о множестве данных \mathbf{X} , из которого берутся выборка \mathbf{X}_n и новые точки, и о механизме извлечения этих данных из множества \mathbf{X} . Другими словами, необходимо определить Модель данных (Data Model), описывающую множество данных $\mathbf{X} \subset \mathbb{R}^p$, и Модель формирования выборки (Sampling Model), описывающую механизм извлечения данных (выборки \mathbf{X}_n и новых данных) из Множества данных \mathbf{X} .

Наиболее популярными моделями данных являются модели данных в виде многообразия (Manifold Data Models), см. [55], [56], [60], [61] и др. работы, для которых множества \mathbf{X} являются q -мерными многообразиями в \mathbb{R}^p , называемыми Многообразиями данных (Data manifold), покрытыми одной координатной системой (картой) [62] и имеющими вид

$$\mathbf{X} = \{X = f(b) \in \mathbb{R}^p: b \in \mathbf{B} \subset \mathbb{R}^q\} \subset \mathbb{R}^p, \quad (17)$$

где $\mathbf{B} \subset \mathbb{R}^q$ - открытое подмножество \mathbb{R}^q , называемое координатным пространством модели, а f - гладкое взаимно-однозначное отображение множества \mathbf{B} в множество $\mathbf{X} = f(\mathbf{B})$. На пару (\mathbf{B}, f)

обычно налагаются различные условия регулярности, основным из которых является наличие у Многообразия данных \mathbf{X} трубки положительного радиуса, обеспечивающее, как следствие, отсутствие у него самопересечений.

Модель формирования выборки определяется обычно как вероятностное пространство $(\mathbf{X}, \sigma(\mathbf{X}), \mu)$ с вероятностной мерой μ на σ -алгебре $\sigma(\mathbf{X})$ измеримых подмножеств \mathbf{X} , носитель $\text{Supp}(\mu)$ которой совпадает с \mathbf{X} . В соответствии с этой моделью, данные \mathbf{X}_n (12) являются повторной выборкой, полученной с помощью вероятностной меры μ на \mathbf{X} .

Можно показать, что модели данных в виде многообразия являются единственными моделями, для которых можно построить решение полной задачи снижения размерности, обладающее свойством состоятельности, означающим, что погрешность в приближенных равенствах (16) стремится к нулю при возрастании объема обучающей выборки n . Если выборка \mathbf{X}_n извлекается из множества данных \mathbf{X} с помощью случайного механизма, то сходимость понимается в вероятностном смысле.

Отображение восстановления g (15) определяет q -мерное эмпирическое многообразие

$$\mathbf{X}_0 = \{X = g(y) \in \mathbb{R}^p : y \in \mathbf{Y}_0 \subset \mathbb{R}^q\} \subset \mathbb{R}^p, \quad (18)$$

в \mathbb{R}^p , покрытое одной картой g с областью определения (координатным пространством) $\mathbf{Y}_0 = h(\mathbf{X})$, а отображение

$$r_0 = g \circ h : \mathbf{X} \subset \mathbb{R}^p \rightarrow \mathbf{X}_0 = r_0(\mathbf{X}) \subset \mathbb{R}^p. \quad (19)$$

отображает многообразие данных \mathbf{X} в эмпирическое многообразие \mathbf{X}_0 .

Соотношения (16) могут быть записаны в виде

$$\mathbf{X} \approx \mathbf{X}_0, \quad (20)$$

понимая под этим выполнение соотношений

$$X \approx r_0(X) \quad \text{для всех } X \in \mathbf{X}. \quad (21)$$

Тем самым, построенное по данным \mathbf{X}_n эмпирическое многообразие \mathbf{X}_0 аппроксимирует (20) многообразие данных \mathbf{X} , и поэтому полная задача снижения размерности в последние годы часто называется задачей моделирования многообразий.

3.3. Снижение размерности в задачах предсказательном метамоделировании

Рассмотрим задачу построения по данным (5) эмпирической зависимости $F_{SM}(X|D_n)$ (4), аппроксимирующей (8) неизвестную функцию $F_M(X)$ (3) в области \mathbf{X} (1) изменения ее аргумента X . Предположим, что множество \mathbf{X} лежит вблизи многообразия \mathbf{M} , размерность q которого меньше исходной размерности p . Пусть по данным \mathbf{X}_n (6), (12) решена полная задача

снижения размерности – построена пара отображений $\theta = (h, g)$ (13), (15), обеспечивающих выполнение приближенных соотношений (16), (21).

Рассмотрим функцию

$$F_{M,\theta}(X) = F_M(r_\theta(X))$$

получающуюся заменой аргумента X функции $F_M(X)$ (3) на значение $r_\theta(X)$ (19), близкое к X (21). Функция $F_{M,\theta}(X)$ зависит от аргумента X только через вектор $y = h(X)$:

$$F_{M,\theta}(X) = f_{M,\theta}(y),$$

где

$$f_{M,\theta}(y) = F_M(g(y)), \quad y \in Y_\theta, \quad (22)$$

а $Y_\theta = h(X)$. Выборка \mathbf{D}_n (5) может быть представлено как множество приближенных значений функции $f_{M,\theta}(y)$ (22):

$$\mathbf{D}_n = \{(y_i = h(X_i), Z_i = f_{M,\theta}(y_i)), i = 1, 2, \dots, n\}, \quad (23)$$

с использованием которой можно быть построена эмпирическая зависимость

$$Z = f_{SM,\theta}(y|\mathbf{D}_n), \quad (24)$$

аппроксимирующая функцию $f_{M,\theta}(y)$ и при достаточном объеме выборки \mathbf{D}_n (23) обеспечивающая выполнение приближенных соотношений

$$f_{SM,\theta}(y|\mathbf{D}_n) \approx f_{M,\theta}(y) \quad \text{для всех } y \in Y_\theta. \quad (25)$$

Эмпирическая зависимость $f_{SM,\theta}(y|\mathbf{D}_n)$ (24) определяет эмпирическую зависимость

$$Z = F^*_{SM}(X|\mathbf{D}_n) = f_{SM,\theta}(h(X)|\mathbf{D}_n),$$

которая в силу соотношения (25) обеспечивает приближенные равенства

$$F^*_{SM}(X|\mathbf{D}_n) \approx F_M(X) \quad \text{для всех } X \in X,$$

аналогичные равенствам (8) для зависимости $F_{SM}(X|\mathbf{D}_n)$ (7). Но для построения зависимости $F_{SM}(X|\mathbf{D}_n)$ необходимо было решать задачу аппроксимации функции, зависящей от p -мерного параметра X , в то время как для построения зависимости $F^*_{SM}(X|\mathbf{D}_n) = f_{SM,\theta}(h(X)|\mathbf{D}_n)$ необходимо было решать редуцированную задачу аппроксимации функции, зависящей от параметра y размерности $q < p$.

Следовательно, использование процедуры снижения размерности позволяет снизить сложность решения основной задачи предсказательного моделирования – построения эмпирической зависимости. Но для возможности такой редукции необходимо, чтобы процедура снижения размерности $\theta = (h, g)$ (13), (15), обеспечивающая выполнение соотношений близости (21), удовлетворяла также условиям функциональной близости

$$F_M(r_\theta(X)) \approx F_M(X) \quad \text{для всех } X \in X.$$

3.4. Снижение размерности в задачах оптимизации

Рассмотрим задачу оптимизации объекта $O \in \mathbf{O}$ с детальным цифровым описанием $X = X(O)$ в соответствии с выбранным критерием оптимизации v , являющуюся формально задачей оптимизации

$$v(X) \rightarrow \text{extr}, \quad X \in \mathbf{X} \subset \mathbb{R}^p \quad (26)$$

в которой \mathbf{X} является q -мерным многообразиями в \mathbb{R}^p вида (17), покрытым одной картой. Если многообразие \mathbf{X} (параметризация $X = f(b)$ многообразия \mathbf{X}) заданы в явном виде, то задача (26) эквивалентна задаче оптимизации $v(b) \equiv v(f(b)) \rightarrow \text{extr}$ функционала $v(b)$ при $b \in \mathbf{B} \subset \mathbb{R}^q$.

Если многообразию \mathbf{X} известно только выборкой \mathbf{X}_n его точек, то множество \mathbf{X} может быть аппроксимировано эмпирическим многообразием \mathbf{X}_0 (17), построенным с помощью решения $\theta = (h, g)$ (13), (15) полной задачи снижения размерности, обеспечивающего выполнение приближенных соотношений (20). Можно ввести репараметризацию координатного пространства \mathbf{B} в представлении (17) многообразия \mathbf{X} таким образом, что многообразию \mathbf{X} может быть записано в виде

$$\mathbf{X} = \{X = f_\theta(y) \in \mathbb{R}^p: y \in \mathbf{Y}_\theta \subset \mathbb{R}^q\} \subset \mathbb{R}^p,$$

в котором координатное пространство $\mathbf{Y}_\theta = h(\mathbf{X})$ совпадает с координатным пространством эмпирического многообразия данных \mathbf{X}_0 (18), а $X = f_\theta(y)$ есть единственное решение уравнения $h(X) = y$ при $y \in \mathbf{Y}_\theta$. Соответственно, задача (26) эквивалентна задаче оптимизации

$$v_\theta(y) \equiv v(f_\theta(y)) \rightarrow \text{extr}, \quad y \in \mathbf{Y}_\theta \subset \mathbb{R}^q. \quad (27)$$

Если процедура снижения размерности θ дополнительно к условиям близости (21) удовлетворяет условию функциональной близости

$$v(X) \approx v(r_\theta(X)) \quad \text{для всех } X \in \mathbf{X}, \quad (28)$$

то вместо исходной задачи оптимизации (26) можно рассмотреть редуцированную задачу оптимизации

$$v_\theta(X) \equiv v(r_\theta(X)) \rightarrow \text{extr}, \quad X \in \mathbf{X} \subset \mathbb{R}^p,$$

которая может быть записана как задача оптимизации

$$V_\theta(y) \equiv v(g(y)) \rightarrow \text{extr}, \quad y \in \mathbf{Y}_\theta \subset \mathbb{R}^q, \quad (29)$$

функционала $V_\theta(y)$, зависящего от q -мерного параметра $y = h(X)$.

Большинство итерационных методов (см., например, [63]) решения задачи оптимизации функционала $v(X)$ обычно используют значения градиента $\nabla_X v(X)$ в различных точках $X \in \mathbf{X}$. Пусть выбран некоторый итерационный метод решения задачи оптимизации (26), использующий значения $\nabla_X v(X)$. При известном параметрическом представлении

многообразия \mathbf{X} выбранный метод определяет соответствующий метод решения задачи (27) и порождает некоторый итерационный оптимизирующий процесс для функционала $v_\theta(y)$ (27) на координатном пространстве \mathbf{Y}_θ .

Применим выбранный метод к редуцированной задаче оптимизации (29). Для того, чтобы этот метод порождал оптимизирующий итерационный процесс для функционала $V_\theta(y)$ (29) на координатном пространстве \mathbf{Y}_θ , близкий к итерационному процессу для функционала $v_\theta(y)$ (27) на том же координатном пространстве, необходимо, чтобы дополнительно к соотношениям (21) и (28) выполнялось дополнительное условие функциональной близости $\nabla_X v(r_\theta(X)) \approx \nabla_X v(X)$ для градиентов функционала v , а также условие тангенциальной близости

$$T(X) \approx T_\theta(r_\theta(X)) \quad \text{для всех } X \in \mathbf{X} \quad (30)$$

касательных подпространств к многообразиям \mathbf{X} и \mathbf{X}_θ в точках $X \in \mathbf{X}$ и $r_\theta(X) \in \mathbf{X}_\theta$ соответственно. Условие (30) эквивалентно условию близости градиентов $\nabla_y v(y) \approx \nabla_y V(y)$ функционалов (27) и (29).

4. Нетрадиционные постановки задач снижения размерности данных

Как было показано в предыдущем разделе, решение полной задачи снижения размерности, которое используется в технологии предсказательного моделирования и оптимизации в качестве промежуточной процедуры, должно удовлетворять дополнительным условиям функциональной и тангенциальной близости. В этом разделе будут даны точные постановки этих нетрадиционных задач снижения размерности.

4.1. Постановка функциональной задачи снижения размерности

Мы сформулируем постановки функциональной задачи снижения размерности для двух различных случаев – при заданных целевых функционалах, близость которых должна обеспечиваться при переходе от исходного к восстановленному цифровому описанию объекта, и для неизвестных функционалов, заданных только выборками их значений.

Задача функционального снижения размерности при известных целевых функционалах.

По выборке \mathbf{X}_n (12) из Множества данных \mathbf{X} , лежащего в пространстве меньшей размерности, построить решение $\theta = (h, g)$ (13), (15), удовлетворяющее условию (16) и дополнительному условию

$$f(g(h(X))) \approx f(X) \quad \text{для всех } X \in \mathbf{X}, f \in \Phi \quad (31)$$

для заданного семейства целевых функционалов Φ , определенных на Множестве данных \mathbf{X} .

Замечание. Некоторое решение этой задачи на базе искусственных нейронных сетей предложено в работах [25], [64].

Задача функционального снижения размерности при целевых функционалах, заданных данными. Пусть целевой векторный функционал

$$Z = f(X) \in R^m, \quad X \in \mathbf{X} \subset R^p, \quad (32)$$

задан множеством своих значений $\{f(X_1), f(X_2), \dots, f(X_n)\} \subset R^m$ на точках выборки \mathbf{X}_n (3) из Множества данных $\mathbf{X} \subset R^p$, лежащего в пространстве меньшей размерности $q < p$. По этим данным требуется построить решение $\theta = (h, g)$ (13), (15), удовлетворяющее условиям (21), (31).

Обычно в постановку задачи функциональной задачи снижения размерности при неизвестных целевых функционалах включают задачу восстановления по данным

$$\mathbf{D}_n = \{(X_1, Z_1 = f(X_1)), (X_2, f(X_2)), \dots, (X_n, f(X_n))\} \quad (33)$$

этих функционалов. Назовем такую расширенную задачу функционального снижения размерности полной и дадим ее формальную постановку.

Полная задача функционального снижения размерности. Для неизвестного целевого функционала f (32), заданного на неизвестном множестве $\mathbf{X} \subset R^p$, лежащем в пространстве меньшей размерности $q < p$, по выборке \mathbf{D}_n (33) построить отображения h (13), g (15) и

$$R_{\text{emp}}: \mathbf{Y} \equiv h(\mathbf{X}) \subset R^q \rightarrow R^m,$$

удовлетворяющие для всех $X \in \mathbf{X}$ условиям (16) и

$$f_{\text{emp}}(X) \equiv R_{\text{emp}}(h(X)) \approx f(X), \quad (34)$$

и для всех $y \in \mathbf{Y}$ условию

$$R_{\text{emp}}(y) \approx R_{\text{emp}}(h(g(y))). \quad (35)$$

Замечание. Из условий (34) и (35) следует, что пара отображений (h, g) удовлетворяет условию (31) в постановке задачи функционального снижения размерности.

4.2. Постановка задачи эффективного снижения размерности

Если из постановки полной задачи функционального снижения размерности исключить требование выполнения условий (16) и (35), и, тем самым, исключить необходимость построения отображения восстановления g , мы приходим к задаче построения по данным \mathbf{D}_n (33) аппроксимирующей зависимости $f_{\text{emp}}(X)$ для неизвестной функции $f(X)$ (32) в предположении, что функция $f(X)$ может быть представлена в виде

$$f(X) = R(h(X)). \quad (36)$$

Тем самым, f является композицией двух отображений: отображения h (13) области определения $\mathbf{X} \subset R^p$ функции $f(X)$ в пространство размерности $q < p$, и отображения

$$R: \mathbf{Y} = h(\mathbf{X}) \subset R^q \rightarrow R^m \quad (37)$$

образа \mathbf{Y} множества \mathbf{X} в \mathbb{R}^m , и функция $f(X)$ зависит от $X \in \mathbb{R}^p$ только через величину $h(X)$ размерности $q < p$. Задачу восстановления функции $f(X)$ в виде функции от аргумента меньшей размерности называют задачей эффективного снижения размерности [65 - 69].

Задача эффективного снижения размерности. Для неизвестной функции $f(X)$ (32), заданной на множестве $\mathbf{X} \subset \mathbb{R}^p$ и имеющей представление (36) в виде композиции отображений h (13) и R (37), якобианы которых имеют полный ранг q , по выборке \mathbf{D}_n (33) построить отображения $h_{\text{emp}}(X)$, $X \in \mathbf{X}$, и $R_{\text{emp}}(y)$, $y \in \mathbf{Y}_{\text{emp}} = h_{\text{emp}}(\mathbf{X})$, такие, что

$$f_{\text{emp}}(X) \equiv R_{\text{emp}}(h_{\text{emp}}(X)) \approx f(X) \quad \text{для всех } X \in \mathbf{X}.$$

Замечание 1. В задаче эффективного снижения размерности не предполагается, что по низко размерному представлению $h_{\text{emp}}(X) \in \mathbb{R}^q$ вектора $X \in \mathbb{R}^p$ можно восстановить этот вектор, но по вектору $h_{\text{emp}}(X)$ можно приближенно восстановить значение $f(X)$ функции (32).

Замечание 2. В цитированных [65 - 69] и других работах, посвященных решению задачи эффективного снижения размерности, всегда предполагается, что отображение h является отображением проектирования на q -мерное линейное подпространство, называемое эффективным пространством задачи.

4.3. Постановка задачи тангенциального снижения размерности

Задача тангенциального снижения размерности является обобщением полной задачи снижения размерности. Предполагается, что множество данных \mathbf{X} является q -мерным многообразием в \mathbb{R}^p , покрытым одной картой и имеющим вид (17). Обозначим $T(X)$ и $T_{\theta}(r_{\theta}(X))$ касательные подпространства к многообразию \mathbf{X} и эмпирическому многообразию \mathbf{X}_{θ} (18), определяемому решением θ полной задачи снижения размерности, в точках $X \in \mathbf{X}$ и $r_{\theta}(X) \in \mathbf{X}_{\theta}$ (19) соответственно. Эти касательные пространства, являющиеся аффинными линейными подпространствами размерности q в \mathbb{R}^p , имеют представления в виде прямых сумм

$$T(X) = X \oplus L(X), \tag{38}$$

$$T_{\theta}(r_{\theta}(X)) = r_{\theta}(X) \oplus L_{\theta}(r_{\theta}(X)), \tag{39}$$

где линейные подпространства $L(X)$ и $L_{\theta}(r_{\theta}(X))$ размерности q в \mathbb{R}^p могут рассматриваться как точки многообразия Грассмана $\text{Grass}(p, q)$, состоящего из q -мерных линейных подпространств в \mathbb{R}^p [70].

Как было сказано в разделе 3.4, для замены сформулированной там задачи оптимизации (26) редуцированной задачей (29) с сохранением структуры оптимизирующих итерационных процессов, необходимы как близость (20) многообразий \mathbf{X} и \mathbf{X}_{θ} , так и близость (30)

касательных подпространств $T(X)$ и $T_\theta(r_\theta(X))$ к ним. В силу представлений (38), (39) и соотношения (21), соотношения (30) эквивалентны соотношениям

$$L_\theta(r_\theta(X)) \approx L(X) \quad \text{для всех } X \in \mathbf{X}, \quad (40)$$

в которых близость в (40) понимается как близость между точками многообразия Грассмана $\text{Grass}(p, q)$ в выбранной метрике на этом многообразии.

Замечание. Как показано в работе [71], близость (40) необходима также для обеспечения обобщающей способности в полной задаче снижения размерности.

Задача тангенциального снижения размерности. По выборке \mathbf{X}_n (12) из неизвестного многообразия данных \mathbf{X} (17) построить решение $\theta = (h, g)$ (13), (15), определяющее эмпирическое многообразие \mathbf{X}_θ (18) и удовлетворяющее условиям (20) и (40).

Заключение.

Рассмотрены основные задачи интеллектуального анализа данных, которые необходимо решать в технологии предсказательного моделирования, основными из которых являются задача построения по данным метамоделей, аппроксимирующих исходные неизвестные модели, и оптимизация построенных метамоделей. При создании метамоделей для ИВС на базе вычислительных экспериментов, проведенных с имитационными моделями ИВС, размерность аргумента метамоделей (параметров, описывающие входные данные) может быть высокой. Однако множество входных параметров для реальных сценариев моделирования имеет внутреннюю размерность, существенно меньшую по сравнению с размерностью исходного аргумента, и может быть аппроксимировано многообразием меньшей размерности, покрытое единственной координатной системой и параметризованное параметром меньшей размерности. Если такое аппроксимирующее параметризованное многообразие построено, то исходные задачи построения метамоделей и их оптимизация могут быть заменены соответствующим редуцированными задачами, зависящими от параметров меньшей размерности.

Задача построения аппроксимирующих многообразий есть задача снижения размерности, называемая также задачей моделирования многообразий. Однако если задача снижения размерности используется в качестве одного из этапов в технологии метамоделирования, а ее решение используется для редуцирования задач построения метамоделей и их оптимизации к задачам, зависящим от параметров меньшей размерности, то решения задач снижения размерности должны удовлетворять дополнительным требованиям, сформулированным в статье. В статье предложены новые нетрадиционные постановки задач снижения размерности с учетом этих дополнительных требований, решения которых могут быть использованы в качестве промежуточных этапов в технологии предсказательного метамоделирования.

Литература.

1. Кулешов А.П. Метамоделли в компьютерных системах проектирования. Труды Международной научно-технической конференции «Информационные технологии и математическое моделирование систем» (ИТММ, 19 – 26 сентября 2010 г., Франция). М.: Учреждение российской академии наук Центр информационных технологий в проектировании РАН, 2010. с. 209-210.
2. Кулешов А.П. Когнитивные технологии в адаптивных моделях сложных объектов. Информационные технологии и вычислительные системы, в. 1, 2008, с. 18 – 29.
3. Кулешов А.П. Технология быстрого вычисления характеристик сложных технических объектов. Информационные технологии, Вып. 3, С. 4-11, 2006.
4. Кулешов А.П. Интеграция данных и знаний при построении метамоделей. Труды Третьей международной конференция «Системный анализ и информационные технологии» (САИТ-2009, 14-18 сентября 2009 г., Звенигород, Россия). 2009, с. 20 – 29.
5. Kuleshov A.P., Bernstein A.V. Cognitive technologies in adaptive models of complex plants. Keynote papers of 13th IFAC Symposium on Information Control Problems in Manufacturing (INCOM'09), June 3 – 5, 2009, Moscow, Russia, p. 70 – 81.
6. Бернштейн А.В., Кулешов А.П. Математические методы построения метамоделей. Труды Третьей международной конференция «Системный анализ и информационные технологии» (САИТ-2009, 14-18 сентября 2009 г., Звенигород, Россия). 2009, с. 756 – 768.
7. Wang, G., Gary, Shan, S. (2007). Review of Metamodeling Techniques in Support of Engineering Design Optimization. *J. Mech. Des.*, 129(3), 370-381.
8. Forrester, A.I.J., Sobester, A., Keane, A.J. (2008). *Engineering Design via Surrogate Modelling. A Practical Guide*. Wiley, New-York.
9. Vapnik, V.N. *Statistical Learning Theory*. Wiley, New-York, 1998.
10. Vapnik, V.N. *The Nature of Statistical Learning Theory*. Springer, New-York, 2000.
11. Mitchel, T. *Machine Learning*. McGraw Hill, New-York, 1997.
12. Бернштейн А.В., Вышинский В.В., Кулешов А.П., Свириденко Ю.Н. Быстрый метод аэродинамического расчета для задач проектирования. Труды Центрального аэрогидродинамического института им. проф. Н.Е. Жуковского. Выпуск № 2678 «Применение искусственных нейронных сетей в задачах прикладной аэродинамики». М.: ЦАГИ, 2008, с. 35- 45.
13. Бернштейн А.В., Вышинский В.В., Кулешов А.П., Свириденко Ю.Н. . Применение искусственных нейронных сетей для определения нагрузок по крылу пассажирского

- самолета на режиме крейсерского полета Труды Центрального аэрогидродинамического института им. проф. Н.Е. Жуковского. Выпуск № 2678 «Применение искусственных нейронных сетей в задачах прикладной аэродинамики». М.: ЦАГИ, 2008, с. 46 – 61.
14. Alexander V. Bernstein Alexander P. Kuleshov, Yuri N. Sviridenko, Victor V. Vyshinsky. Fast Aerodynamic Model for Design Technology/ Proceedings of West-East High Speed Flow Field Conference. Nov. 19-22, 2007, Moscow, Russia, <http://wehsff.imamod.ru/pages/s7.htm>.
 15. Мизин И. А., Богатырев В. А., Кулешов А. П. Сети коммутации пакетов/ Под ред. В. С. Семенихина. — М.: Радио и связь, 1986. — 408 с.
 16. Кулешов А.П., Бернштейн А.В. Предсказательное метамоделирование и интеллектуальный анализ данных. Двенадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2010). Труды конференции. - Т. 1. - М.: Физматлит, 2010. - с. 192 - 200.
 17. Бернштейн А.В., Бурнаев Е.В., Кулешов А.П. Интеллектуальный анализ данных в метамоделировании. Нейроинформатика, ее приложения и анализ данных. Материалы XVII Всероссийского семинара (2 – 4 октября 2009 г.), Красноярск, ИПК СФУ, 2009, с. 23 – 28.
 18. Бернштейн А.В., Кулешов А.П. Когнитивные технологии интеграции данных и знаний в метамоделировании. Обзорные прикладной и промышленной математики, серия «Вероятность и статистика». 2009, т. 16, № 6, с. 1088 – 1089.
 19. Bernstein A.V., Kuleshov A.P. Computer Data Analysis in Metamodeling. Proceedings of the 9th International Conference on Computer Data Analysis and Modeling: Complex Stochastic Data and Systems (CDAM'2010), September, 7-11, 2010, Minsk, Belarus, 2010.
 20. Kuleshov A.P., Bernstein A.V. Predictive Modeling and Data Analysis. Proceedings of the 8th International Conference on Intelligent Information Processing (IIP-2010), October, 17-24, 2010, Republic of Cyprus, Paphos, M.: МАКС Пресс, pp. 212 – 215.
 21. Кулешов А.П., Бернштейн А.В. Проблемы интеллектуального анализа данных для создания метамоделей. Труды Международной научно-технической конференции «Информационные технологии и математическое моделирование систем» (ИТММ, 19 – 26 сентября 2010 г., Франция). М.: Учреждение российской академии наук Центр информационных технологий в проектировании РАН, 2010. С. 222-223.
 22. Кулешов А.П. Задачи многомерного статистического анализа в системах компьютерного проектирования. Труды VII Международной школы-семинара «Многомерный статистический анализ и эконометрика» (Армения, Цахкадзор, 2008), М., 2008, 60 – 61.

23. Бернштейн А.В. Интеллектуальный анализ данных в теории надежности. Сборник докладов 6-й Международной конференции по Математическим методам теории надежности (MMR-2009), Москва, 22-27 июня 2009г., с. 109 – 113.
24. Бернштейн А.В., Кулешов А.П. Построение ортогональных нелинейных многообразий в задачах снижения размерности. Труды VII Международной школы-семинара «Многомерный статистический анализ и эконометрика» Цахкадзор, 2008, 25 – 27.
25. Бернштейн А.В., Кулешов А.П. Когнитивные технологии в проблеме снижения размерности описания геометрических объектов/ Информационные технологии и вычислительные системы. 2008, №2, с. 6 – 19.
26. Иванова Е.П., Чернова С.С. Снижение размерности сложных геометрических объектов при наличии частных параметрических моделей. Искусственный интеллект и принятие решений, 2009, №3, с. 53 - 58.
27. Бурнаев Е.В., Чернова С.С., Сартори А.В., Свириденко Ю.Н. О методах консолидации многомерных данных. Труды Третьей международной конференция «Системный анализ и информационные технологии» (САИТ-2009, 14-18 сентября 2009 г., Звенигород, Россия). 2009, с. 789 – 799.
28. Бурнаев Е.В., Чернова С.С., Беляев М.Г., Свириденко Ю.Н. О методах построения консолидированных данных. Обзорение прикладной и промышленной математики, серия «Вероятность и статистика». 2009, т. 16, № 6, 1035 - 1036.
29. Бурнаев Е.В., Бернштейн А.В. Методы консолидации разноточных данных. Труды 8-й международной конференции «Интеллектуализация обработки информации» (ИОИ-2010), 17- 24 октября 2010, Кипр. М.: МАКС Пресс, 2010. С. 220-223.
30. Бочаров П.П., Печинкин А.В. Теория массового обслуживания. М.: Изд-во РУДН, 1995 г.
31. Клейнрок Л. Теория массового обслуживания. М.: Изд - во Машиностроение, 1979 г.
32. Ивченко Г.И., Каштанов В.А., Коваленко И.Н. Теория массового обслуживания. М.: Изд – во Высшая школа, 1982 г..
33. Саати Т.Л. Элементы теории массового обслуживания и ее приложения. М.: Изд – во Советское радио; 1965 г.
34. Sacks, J., Welch, W.J. Design for Computer Experiments, Technometrics, 1989, 31(1), 41-47.
35. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn H.P. Design and Analysis of Computer Experiments, Statistical Science, 1989, 4(4), 409-435.
36. Martin, J.D., Simpson T.W. Use of Kriging Models to Approximate Deterministic Computer Models. AIAA Journal, 2005, 43(4), 853-863.

37. Cressie, N. Spatial Prediction and Ordinary Kriging. *Mathematical Geology*, 1988, 20(4), 405-421.
38. Hagan M. T. and Menhaj M. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, Vol:5, No:6, pp.989 - 993, 1994.
39. Dorofeyev Ye.A., Sviridenko Yu.N. The application of replicator neural networks to the problems of aerodynamic design of aircraft components//International Symposium on Aerospace Technologies of the XXI century: New Challenges in Aeronautics. Zhukovsky, Russia, 14 – 19 August 2001.
40. Bernstein A.V., Burnaev E.V., Chernova S.S., F. Zhu, N. Qin. Comparison of Three Geometric Parameterization methods and Their Effect on Aerodynamic Optimization. *Proceedings of International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems (Eurogen 2011)*. Capua, Italy, September 14 – 16, 2011.
41. Кашафутдинов С.Т., Лушин В.Н. Атлас аэродинамических характеристик крыловых профилей, Новосибирск, 1994.
42. T.F. Cox, M.A.A. Cox. *Multidimensional Scaling*. Chapman and Hall, 2001.
43. T. Jollie. *Principal Component Analysis*. Springer, New-York, 2002.
44. T. Hastie, R. Tibshirani, J. Friedman. *The elements of Statistical learning*. Springer, New-York, 2001.
45. P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, p. 711 – 720.
46. M.A. Turk, A.P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.
47. Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, Hong-Jiang Zhang. Face Recognition Using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3), 2005, p. 328 – 340.
48. Talwalkar Ameet, Kumar Sanjiv, Rowley Henry. Large-Scale Manifold Learning. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, [CVPR 2008 papers on the web \(http://www.cvpapers.com/cvpr2008.html\)](http://www.cvpapers.com/cvpr2008.html).
49. L.K. Saul, S.T. Roweis. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, 2000, p. 2323 – 2326.
50. J.B. Tehenbaum, V. de Silva, J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2000, p. 2319 – 2323.

51. M. Belkin, P. Niyogi. Laplacian eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Processing Systems (NIPS)*, MIT Press, v. 14, 2002.
52. M. Belkin, P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15, 2003, p. 1373 – 1396.
53. D.L. Donoho, C. Grimes. Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. Technical Report 2003-08, Stanford Statistical Department, 2002.
54. K. Q. Weinberger, L. K. Saul. Maximum Variance Unfolding: Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1), 2006, p. 77 – 90.
55. Zhenyue Zhang and Hongyuan Zha, Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment, *SIAM Journal on Scientific Computing* 26 (1) (2005), 313–338.
56. De-Li Zhao. Tangential Eigenmaps: A Unifying Geometric for Manifold Learning. Shanghai Jiao Tong University, Oct. 2005 (Unpublished: see <http://sites.google.com/site/zhaodeli/paper>).
57. Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, M. Ouimet. Learning Eigenfunctions Link Spectral Embedding and Kernel PCA. *Neural Computation*, 16(10), 2004, p. 2197 – 2219.
58. Y. Bengio, J.-F. Paiement, P. Vincent. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. *Advances of Neural Information Processes Systems (NIPS)*, Volume 15, 2003.
59. Anders Brun, Carl-Fredrik Westin, Magnus Herberthson, Hans Knutsson. Fast Manifold Learning Based on Riemannian Normal Coordinates. *Proceedings of the 14th Scandinavian conference on image analysis (SCIA'05)*, Joensuu, Finland, June, 2005, pp. 920 - 929.
60. Lawrence Cayton. Algorithms for manifold learning. Spring 2005, UCSD tech report CS2008-0923.
61. M. Brand. Charting a manifold. *Advances in Neural Information Processing Systems*, 15, MIT Press, 2003.
62. Бишоп Р., Криттенден Р. Геометрия многообразий. М., Мир, 1967.
63. Nesterov Yu. *Introductory Lectures on Convex Programming*. Boston: Rluwer, 2004.
64. Бернштейн А.В., Кулешов А.П. Задачи снижения размерности моделей сложных объектов/ Вторая Международная конференция «Системный анализ и информационные технологии» САИТ-2007 (10 – 14 сентября 2007 г., г. Обнинск, Россия): Труды конференции. Т. 1, стр. 243 – 247.

65. Li, K. C. Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 1991, 86, 316–342.
66. Li, K. C. On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. *Journal of the American Statistical Association*, 1992, 87, 1025–1039.
67. Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X. An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Ser. B*, 2002, 64, 363–410.
68. Carroll, R. and Li, K. C. Measurement error regression with unknown link: dimension reduction and data visualization. *Journal of the American Statistical Association*, 1992, 87, 1040–1050.
69. Wang Qin, Yao Weixin. An adaptive estimation of MAVE. *J. Multivariate Analysis*, 104(1), 2012, p. 88 – 100.
70. R.P. Woods. Differential geometry of Grassmann manifolds. *Proc. Nqt. Acad. Sci. USA*, 57 (1967), p. 589 – 594.
71. Bernstein A.V. Local generalizing ability in Manifold Learning Problem. Proceedings of the International conference “Statistical Models and Methods for Reliability and Survival Analysis and Their Validation” (S2MRSA’ 2012), Bordeaux, July 4 - 6, 2012.

Агалаков Юрий Глебович. Генеральный директор ОАО НИИ автоматической аппаратуры им. акад. Семенихина. Окончил физфак ЛГУ в 1986 г., кандидат физико-математических наук, автор более 30 научных работ. Область научных интересов: информационные и телекоммуникационные технологии, моделирование.

Бернштейн Александр Владимирович. Заведующий лабораторией Когнитивных методов анализа данных и моделирования Института Системного Анализа РАН. Окончил механико-математический факультет МГУ им. Ломоносова в 1969 г. Доктор физико-математических наук, профессор. Автор более 100 печатных научных работ. Область научных интересов: теоретическая и прикладная математическая статистика, интеллектуальный анализ данных, моделирование.